

UNITED STATES PATENT APPLICATION

**METHOD AND APPARATUS FOR TRANSMITTING HIGH SPEED DATA BY
SPECTRAL DECOMPOSITION OF THE SIGNALING SPACE**

Inventor

Adrian S. Matthews

Cross Reference to Related Application

This is a continuation-in-part of a pending United States patent application Serial number 09/021,137 filed February 10, 1998

METHOD AND APPARATUS FOR TRANSMITTING HIGH SPEED DATA BY SPECTRAL DECOMPOSITION OF THE SIGNALING SPACE

Field of the Invention

[0001] This invention is generally directed to high-speed data communication, and more specifically, to the area of high-speed modem design. It relates to achieving high spectral efficiency in signaling systems.

Background of the Invention

[0002] Modern telecommunication applications have resulted in substantial increases in the need for additional bandwidth. For example, in the area of wired communications, there is a need to simultaneously support voice, video, and data applications at low BER (Bit Error Rates) using new high-speed modem designs for twisted pairs. At signaling rates better than 10Mbits/s performance bounds generally exceed a BER of 10^{-6} . When the conventional Pulse Amplitude Modulation (PAM) technique is used, the baseband communication signal is represented by a series of modulated pulses whose amplitude levels are determined by the symbol to be transmitted. For example, with 16-QAM (Quadrature Amplitude Modulation), typical symbol amplitudes of ± 1 and ± 3 are utilized in each quadrature channel. For digital communication systems, efficient use of bandwidth is crucial when dealing with time-dispersive channel, as is common with wireless systems. In these types of systems, whenever there is distortion of the signals due to preceding or following pulses, normally referred to as pre-cursors and post-cursors, respectively, the amplitude of the desired pulse is affected due to superimposition of the overlapping pulses. This phenomenon is known as intersymbol interference, and is an impediment to high-speed data transmission, especially in systems that are constrained by limited bandwidth.

[0003] One way to minimize the effects of intersymbol interference is to use an equalizer. Fixed equalizers are designed to be effectively operated between an upper and lower bound between which the channel is expected to deviate. Whenever these limits are exceeded, the equalizer ceases to operate effectively. Hence, there has to be a certain degree of precision when channel equalization is employed, and fixed

equalizers are implemented. There are adaptive equalizers (i.e., continuous) that track dynamic channel dispersion and make continuous adjustments to compensate for such intersymbol interference. This provides some improvement in performance over the fixed equalizer.

[0004] Incorporation of the equalizer into some communication systems does not come without penalty. In wireless systems, for instance, insertion loss becomes a critical factor if the equalizer is present and the associated impairment does not occur. The main purpose of the equalizer implementation is to enhance the information bearing capability of the communication system with the design objective of asymptotically approaching the capacity bounds of the transmission channel. Consequently, the use of the equalizer can be regarded as one instance of an array of possibilities that may be implemented to enhance the bit rate of the communication system design.

Summary of the Invention

[0005] In accord with the invention, a method and apparatus is provided that makes more efficient use of the available signaling bandwidth in the sense of asymptotically approaching possible transmission limits. This is done by significantly reducing the effects of intersymbol and interchannel interference by a judicious choice of the signaling pulse shapes. In particular, prolate pulses are used to extend channel capacity and reduce interference. By use of orthogonal axes that span the signal space, combined with water filling techniques for efficient allocation of transmission energy based on the noise distribution, the information content can be increased without increase in bandwidth.

[0006] The signaling space is spectrally decomposed to support the simultaneous transmission of multiple signals each with differing information bearing content. These signals being orthogonal, are non-interfering. Signals are constructed as complex sets and are generally represented with axial coordinates, all orthogonal to one another within the complex plane. The real axis is termed the in-phase (I) component and the imaginary axis is termed the quadrature (Q) component. Each component defines a spanning vector in the signal space.

Brief Description of the Drawings

[0007] FIG. 1 is a graph depicting the concentration of energy in a prolate pulse interval;

[0008] FIG. 2 is a block schematic illustrating an optimized modem using the invention;

[0009] FIG. 3 is a block schematic of a Single Segment Discrete Prolate Transmitter;

[0010] FIG. 4 is a graph illustrating the application of water filling to the present invention.

[0011] FIG. 5 is a graph depicting channel segmentation and use of the frequency response; and

[0012] FIG. 6 is a block schematic of the Discrete Prolate demodulator corresponding to FIG. 3.

Detailed Description

[0013] Spectral efficiency in digital systems is largely a function of the wave shapes of the signals that are used to carry the digital information. There are tradeoffs between time limitations and frequency limitations. These two requirements generally have a flexible relationship. The characteristics of prolate pulses may be chosen to limit spectral energy dispersion thereby permitting more signaling channels for a given bandwidth. These advantages become readily apparent with an analysis of the prolate pulse spectral performance. In particular, the Fourier transform of the waveform is very band limited. Proper selection of signal space such as axes or spectral vectors representing signal coordinates are very important. If signals are orthogonal to one another, transmission techniques utilizing methods of water filling may be implemented with significant increase in efficiency. The technique of water filling is discussed in "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come", J. A. C. Bingham, "IEEE Communications," May 1990, pp. 5 – 14 incorporated by reference. An illustration of water filling may be ascertained from the graph of FIG. 4. A bandwidth of a channel is defined by the marks 401 and 402 on the horizontal axis. The curve 403 defines the noise level produced as seen by a receiver. The energy level, which the

channel can transmit, is defined by the horizontal level 404. The area 405 bounded by the curve 403 and energy level 404 may be “water filled” by data signals. The data acceptance area 405 of the band is divided into sections 408 by vertical dividers 409. The signal data is inserted into a section until the added data and noise in that section reaches the energy level limit. This filling combined with the orthogonal nature of the data signals inserted in the sections permit the increase in the data capacity of the channel.

[0014] Consider a trigonometric polynomial $p_i(t)$ defined as follows:

$$p_i(t) = \sum_{n=-N}^N a_n e^{jn\pi} \quad (1)$$

[0015] In equation (1) the period may be chosen to be 2 by suitable scaling of t . The coefficients a_n can be obtained by an optimization process, the objective of which is to obtain a spectrally efficient pulse. The process may be regarded as a scheme in which the energy of the pulse is concentrated in the interval $[-\varepsilon, \varepsilon]$. This is shown in Figure 1 where a more or less generic pulse 101 is shown and the constraining interval 102 is indicated. The optimization process is a transmission pulse design problem, and a particular mathematical approach for achieving this objective is now described. In general, optimal communication system design requirements often necessitate the transmission of spectrally efficient pulses in order to minimize both intersymbol interference and interchannel interference where application requires segmented spectrum utilization.

[0016] Based on the specified format in equation (1), it can be shown that the coefficients a_n of $p_i(t)$ satisfy the following system of equations:

$$\sum_{m=-N}^N \frac{\sin(n-m)\pi\varepsilon}{(n-m)\varepsilon} a_m = \lambda a_n, \quad n = -N, -N+1, \dots, N. \quad (2)$$

[0017] Equation (2) may be rewritten in the form,

$$S\bar{a}_i = \lambda_i \bar{a}_i \quad (3)$$

[0018] Where the coefficients S_{nm} of the matrix S defined by equation (3), and eigenvectors \bar{a}_i are given by,

$$S_{nm} = \frac{\sin(n-m)\pi\epsilon}{(n-m)\epsilon} \quad (4)$$

and

$$\vec{a}_i = [a_{-Nt}, a_{(-N+1)t}, \dots, a_{0t}, \dots, a_{(N-1)t}, a_{Nt}]' \quad (5)$$

[0019] Where t denotes transpose. The matrix S is a real, symmetric, and positive definite with other mathematical properties of interest to the development, as now discussed. There are thus $2N+1$ real eigenvalues λ_i which satisfy (3) and which may be ordered such that:

$$\lambda_1 > \lambda_2 > \dots > \lambda_{2N+1} \quad (6)$$

[0020] For each eigenvalue λ_i , there is an associated eigenvector \vec{a}_i , whose coefficients can be used to form the trigonometric function defined in equation (1). The eigenvectors of the matrix S may be normalized to have unit energy. Because of the orthogonality of the eigenvectors of symmetric matrices, their dot products $\vec{a}_i \bullet \vec{a}_j$ satisfy the following relationship,

$$\vec{a}_i \bullet \vec{a}_j = \sum_{n=-N}^N a_{in} a_{jn} = \delta_{ij}, \quad (7)$$

[0021] Where δ_{ij} is the Kronecker delta function. Because of equation (3) and equation (7), it can be shown that functions of the form of equation (1) whose coefficients are those of the eigenvectors of the matrix S as defined in equation (4), the following relationships holds:

$$\frac{1}{2} \int_{-1}^1 p_i(t) p_j(t) dt = \delta_{ij}, \quad (8)$$

and,

$$\frac{1}{2} \int_{-\epsilon}^{\epsilon} p_i(t) p_j(t) dt = \lambda_i \delta_{ij} \quad (9)$$

Functions so formed are described as discrete prolate.

[0022] With the background material discussed above, a particular method of communicating digital information using the functions $p_i(t)$ defined earlier is now

presented. Again, in view of equation (6), there are $2N+1$ eigenvectors that satisfy equation (3). The vectors together form a spanning set for the vector space defined by the matrix S . Define D to be the dimension of the associated vector space. Then D is given by:

$$D = 2N + 1 \quad (10)$$

[0023] Note that D is a design parameter, and is a function of N . By analogy, $\{p_i(t)\}$ form a spanning set for the signal space associated with the matrix S , and this signal space is also D dimensional. Consider the construct:

$$x_i(t) = \sum_{k=-\infty}^{\infty} I_k p_i(t - kT) \quad (11)$$

[0024] Generalizing and using equation (8), it can be shown that the following holds:

$$\frac{1}{2T} \int_{-T+kT}^{T+kT} x_i(t) p_j(t) dt = \delta_{ij} I_k \quad (12)$$

[0025] Equation (12) is of critical importance to the invention. The implications are that if a function of the form of equation (11), for a specific value of i , is transmitted over a communication channel, then the transmitted alphabet I_k will only be uniquely determined in an interval defined by k if the corresponding $p_i(t)$ is used as the receiving filter. If a function of the form of equation (11), for a specific value of i , is transmitted over a communication channel, and $p_j(t)$ for $j \neq i$, is used as the receiving filter, then the function $p_i(t)$ will be virtually non-existent. Thus in order to extract the information content of a signal whose format is given by equation (11), the signaling pulse must be matched at the receiver. In anticipation of making reference to Cartesian space, the format of equation (11) is used in the construction of $y_i(t)$ defined as follows:

$$y_i(t) = \sum_{k=-\infty}^{\infty} Q_k p_i(t - kT) \quad (13)$$

where again Q_k is the alphabet to be transmitted. It is clear that equation (13) also satisfies a relationship similar to equation (12). Equations (11) and (13) can now be

used to quadrature modulate a carrier in the final part of the transmission signal synthesis. Define $s_i(t)$ by:

$$s_i(t) = x_i(t) \cos(2\pi f_c t) - y_i(t) \sin(2\pi f_c t) \quad (14)$$

[0026] Thus, the signals are constructed as complex sets and are generally represented as vectors within the complex plane. The real axis is termed the in-phase (I) component and the imaginary axis is termed the quadrature (Q) component.

[0027] As indicated by equation (10), there are D such constructs possible. Because of the Orthogonality of the building blocks $\{p_i(t)\}$ discussed earlier, $\{s_i(t)\}$, being linear combinations of a single $p_i(t)$ for each i , are themselves orthogonal, forming a spanning set for the signal space defined over the channel band limited by $W = 1/2T$. That is to say, each such signal $s_i(t)$ may be regarded as an orthogonal “finger” over which the symbols $\{I_k, Q_k\}$ may be independently transmitted. Thus, equation (14) can be used to increase the bit rate of the communication channel without bandwidth expansion. Of course coding and equalization may be added to improve fidelity.

[0028] The parameters ε and N determine the spectral shape of the transmission pulses $p_i(t)$. In general ε will be used to determine the compactness of the fit within the signaling period, while N determines the peaking and roll-off. It is important for N to be fairly large ($N \geq 10$) as there are at least two benefits to be gained in this regard. Firstly, large values of N contribute to better roll-off characteristics, which directly minimize intersymbol interference. Secondly, as can be seen by equation (10), large values of N contribute directly to an increase in the dimension of the signaling space, providing more discrete prolate functions that can be used to increase the capacity of the transmission system design. However, these benefits must be balanced by the fact that tighter peaks that are made possible by larger values of N are likely to place greater implementation constraints on the receiver, to the extent that more accurate symbol timing shall be required to retrieve the encoded digital information.

[0029] In general, the range of values that can be taken on by the discrete symbols $\{I_k, Q_k\}$ determines the number of levels M that may be reasonably distinguished at the receiver, with noise, crosstalk, and interference playing a critical role in the process. Conventional modulation techniques, such as QAM for instance, may be referenced, and the value of M shall be determined in an optimization process in which the power is held constant, and the bit rate is maximized for a given BER constraint. Given M , the information bearing capacity C of the transmitter is computed in a straightforward manner. Thus,

$$C = \frac{\log_2 M}{T} \quad (15)$$

[0030] Where C is given in units of bits/s. Equation (15) holds for the one-dimensional case. That is, only when one signal of duration T having M reasonably distinguishable levels is transmitted in a channel bandlimited by W . However, when multiple orthogonal signaling is used for data transmission, the parameter M in equation (15) will be a function of the number of signals chosen, along with the associated levels that may be represented by each independent choice. A limiting reformulation of equation (15) is now given by:

$$C_{\text{lim}} = \frac{\log_2 \prod_i M_i}{T} \quad (16)$$

[0031] Equation (16) hints of the possibility of channel optimization with more efficient encoding of the signaling data. A concrete use of equation (16) is demonstrated in the sequel, specifically, with the aid of FIGs. 3 and 6.

[0032] The area of optimal communication system design is generally one in which various signal-processing techniques are comprised to asymptotically approach theoretically established channel capacity limits. Transmission rates may further be optimized if a process known as water filling is implemented. With the implementation of water filling, the available signaling power is allocated to the communication channel, and the bits are loaded in a manner related to the noise spectral density, with the objective of maximizing resource utility. It may be regarded as a process in which the squander of the available signal energy is avoided. Let the noise be Gaussian, with the

power spectral density given by $N(f)$, with $H(f)$ being the associated complex transfer function of the channel. Then, in order to make efficient use of the available signaling power S , the optimal channel input power is given by:

$$S = \int_{f \in \Omega} B - \frac{N(f)}{|H(f)|^2} df \quad (17)$$

where the region of integration Ω is defined by:

$$\Omega = \left\{ f : \frac{N(f)}{|H(f)|^2} \leq B \right\} \quad (18)$$

[0033] Equations (17) and (18) describe the elastic relationship that exists between the power spectral densities of the input signal and the noise during the process of optimizing the available channel bandwidth. In equations (17) and (18) B is an average input power constraint.

[0034] From a practical standpoint, the optimal allocation of signaling power is best achieved by channel segmentation. Then, the allocation of bits to the various sub channels is achieved through the process of maximizing the channel capacity while minimizing the baud error rate. There exist in the literature a variety of optimal loading algorithms through which the required energy distribution may be accomplished. A good example may be found in patents: "Ensemble Modem Structure For Imperfect Transmission Media" US patents 4,679,227, 4,732,826 and 4,833,706.

[0035] With the aid of FIG. 5, an exemplary approach to optimized loading is now discussed. Let the available channel bandwidth be divided into N equal segments of length W . Assume that the frequency response within the i^{th} segment is flat and given by $H_i(f)$. Let the noise be additive white Gaussian with double-sided spectral density $N_o/2$ watts/Hz. Let the available signal power P be equally divided among all sub channels available, and normalize the system to the first sub channel. The received power in the i^{th} sub channel is thus $P_i = l_i P / N$ where $l_i = |H_i(f)|^2 / |H_1(f)|^2$. It can then be shown that a possible optimal choice of bit loading n_i is given by:

$$n_i = \log_2 \left\{ \frac{(l_i / N)(3/2)(P / N_o W)}{-\ln \Pr(\varepsilon)} \right\} \quad (19)$$

[0036] Where n_i is the number of bits allocated to the i^{th} sub channel, and $\Pr(\varepsilon)$ is the probability of symbol error for all sub channels. In equation (19), since $I_i P / N_o W$ is the signal to noise ratio in the i^{th} sub channel, a preferred embodiment of the invention will use a measured value of the noise in the i^{th} sub channel for the computation of n_i . This combined approach to the allocation of signaling energy and of bits to each sub channel comprises a specific optimal approach to water filling.

[0037] A block diagram of the complete transmitter/receiver pair is shown in FIG. 2. In FIG. 2 the transmitter comprises N sub-transmitters 201-1 to 201-N and the summer 202. Input data for transmission through the channel are modulated at each sub-transmitter 201-1 to 201-N, and the outputs are summed at the summer 202 for transmission through the channel characterized by the function block 203. Addition of noise into the system is depicted by function block 204 in FIG. 2. The i^{th} sub-transmitter 201- i is optimized in accord with water filling as described above for the i^{th} segment of the channel. Similarly, the receiver is comprised of N subcomponents 205-1 to 205-N, the i^{th} subcomponent 205- i corresponding the component 201- i of the transmitter.

[0038] The invention is now further described with greater specificity with the use of FIG. 3, which illustrates how the discrete prolate functions are used for capacity optimization. With reference to function block 201-1, assume that, with the use of equation (19), a computed value of 6 was obtained for n_1 . It is clear from the foregoing discussion that this loading bound can be assured through the resolution of the transmitted signal with the use of two discrete prolate functions. Let $n_1 = n_{11} + n_{12}$ with $n_{11} = 2$ and $n_{12} = 4$. It can further be shown that, given the specific choices for n_{11} and n_{12} , if it is assumed that the symbol error is equal in both signaling dimensions, the power must be divided such that $P_1 = P_{11} + P_{12}$, where $P_{11} = P_1 / 3$ and $P_{12} = 2P_{11}$. Given the foregoing choices of parameters, an exemplary embodiment of the invention in function block 201-1 is illustrated in FIG. 3. As can be seen from the figure, the six bits to be transmitted are segmented at function block 301 into 2- and 4-bits packets that are sent to function blocks 302-1 and 302-2. At function block 303-1, a 4-level I/Q mapper is

used, while a 16-level mapper is used at function block 303-2. Within function block 304-1 the I and Q components from function block 303-1 and the power P_{11} are used to generate the in-phase and quadrature components of the prolate pulses corresponding to $p_1(t)$. Further, these components are modulated at function block 306-1 and 308-1, then summed at function block 309 for output to the channel. As can be seen from FIG. 3, similar activities occur for the dimension corresponding to $p_2(t)$.

[0039] The structure of the optimized sub-receiver 205-1, associated with sub-transmitter 201-1, is shown in FIG. 6. As discussed earlier, the key to retrieving the bits that were sent in a particular dimension is the use of a low pass eigenfilter for that dimension. The discrete prolate pulses are thus used to form a low pass orthogonal filter bank for extracting the bit information from each dimension. The demodulated I_k and Q_k values finally go through a reverse mapping process, after which the original block of bits is reconstructed.

[0040] In the receiver of FIG. 6 the channel output is received as indicated by the block 601. This channel output is connected to a plurality of mixers 603-1, 603-2, 605-1 and 605-2 and are mixed with cosine and sine signals, respectively. These mixed signals are demodulated in the orthogonal filter bank containing filters 606-1, 606-2, 606-3 and 606-4. I/Q reverse mappings are performed in reverse mappers 607 and 608 to recover the segmented bits and the originally transmitted bit pattern is reconstructed in block 609. While discrete blocks are illustrated, the processes are stored program processes that are performed independently of block identification.

[0041] Synchronousness being of critical significance to the design of telecommunication systems, reference is now made to the fact that in the construction of FIG. 2, FIG. 3, and FIG. 6, this requirement is stipulated. Thus, in a complete embodiment of the present invention, methods of carrier tracking and symbol recovery shall be implemented. There are various procedures well documented in the literature to accomplish these operations. One reference describing synchronism with respect to carrier tracking and symbol recovery is the text "Digital Communications, Fundamentals

and Applications” by Bernard Sklar. Information specifically related to synchronization may be found in chapter 8, Pages 429-474.

[0042] Recall that in equation (2) ε was used to determine the pulse efficiency. Thus, in a preferred embodiment equation (9) may be used to shorten the length of the filtering process, in an effort to seek implementation efficiency. Filtering must then be normalized by a factor of $1/\lambda_i$ for each finger. In this case, keeping jitter to a minimum will be a critical issue.

[0043] In present-day communication systems, because of the inefficiencies that occur with the application of a single signal for information bearing, the implementation of complex equalization structures is imperative to achieve the most efficient use of the channel. With the implementation of the design discussed herein, the equalizer shall effectively be reduced to a simple scaling function.

[0044] The invention presented herein was described in light of a preferred embodiment. It should be understood that such preferred embodiment does not limit the application of the present invention. Persons skilled in the art will undoubtedly be able to anticipate alternatives that are deemed to fall within the scope and spirit of the present application.